



**ITESO**

**Universidad Jesuita  
de Guadalajara**

Proyecto 004 Pairs Trading

Materia microestructura y sistemas de trading

Alumna Keysi Fuentes Piña

Profesor Luis Felipe Gómez Estrada

Otoño 2025

## Project Overview

Este proyecto desarrolla una estrategia completa de Statistical Arbitrage basada en Pairs Trading, implementada mediante:

- Identificación de pares cointegrados
- Estimación del hedge ratio dinámico con Filtros de Kalman
- Generación de señales mediante Johansen + VECM (Vector Error Correction Model)
- Evaluación en un entorno realista con walk-forward (Train–Test–Validation)

El filtro de Kalman se formula explícitamente como un Sequential Decision Process siguiendo el marco de Powell:

predict → observe → update → decide → act → learn.

## Objectives

- Identificar pares cointegrados mediante pruebas estadísticas
- Implementar dos filtros de Kalman como modelos de decisión secuencial
- Construir una estrategia 100% market-neutral
- Analizar desempeño considerando costos reales
- 0.125% comisión por trade
- 0.25% tasa de financiamiento anual
- Crear código modular y limpio
- Generar todas las métricas profesionales de rendimiento

## Strategy Description and Rationale

La estrategia desarrollada se basa en Pairs Trading, una técnica clásica de statistical arbitrage que busca explotar desviaciones temporales entre dos activos que mantienen una relación de equilibrio en el largo plazo. A diferencia de estrategias direccionales, Pairs Trading es market-neutral, lo que significa que su comportamiento no depende de si el mercado sube o baja, sino de la relación entre los dos activos seleccionados.

¿En qué consiste Pairs Trading?

Se seleccionan dos activos cuyos precios mantienen una relación estable, expresada normalmente como:

$$Y_t - \beta X_t = \epsilon_t$$

donde:

- $\beta$  es el hedge ratio,
- $\epsilon_t$  es el spread o residual.

Si  $\epsilon_t$  es estacionario ( $I(0)$ ), la relación está cointegrada.

Cuando el spread se desvía demasiado de su promedio, se espera que eventualmente regrese (mean reversion).

### **Key Concepts**

Cointegration

Relación de equilibrio de largo plazo:

$$Y_t - \beta X_t = \epsilon_t, \epsilon_t \sim I(0)$$

Si el residual es estacionario  $\rightarrow$  el spread es mean-reverting.

Mean Reversion

Cuando el spread se aleja demasiado del equilibrio, se espera que regrese  $\rightarrow$  oportunidad de arbitraje.

Dynamic Hedging (Kalman Filter)

$$\begin{aligned} x_t &= Fx_{t-1} + w_t \\ y_t &= H_t x_t + v_t \end{aligned}$$

Permite estimar  $\alpha_t, \beta_t$  en tiempo real.

VECM

Modelo de cointegración:

$$ECT_t = \beta_x X_t + \beta_y Y_t$$

Sequential Decision Process (Powell)

1. Estado
2. Información
3. Transición
4. Observación

5. Decisión
6. Aprendizaje

Ambos Kalman + estrategia son SDA completos.

### Market Neutral Strategy

Exposición sistemática ≈ 0

80% del portafolio invertido:

40% en cada activo.

### Technical Requirements

#### Data Requirements

-15 años de precios diarios

Tickers seleccionados: **AAPL, MSFT, NVDA**

 AAPL_prices_20100101-20250101	13/11/2025 01:10 p. m.	Archivo de valores...	117 KB
 historical_prices_20100101-20250101_3ti...	13/11/2025 01:10 p. m.	Archivo de valores...	258 KB
 MSFT_prices_20100101-20250101	13/11/2025 01:10 p. m.	Archivo de valores...	116 KB
 NVDA_prices_20100101-20250101	13/11/2025 01:10 p. m.	Archivo de valores...	118 KB

Limpieza y manejo de NA

Split:

60% Train

20% Test

20% Validation

### Transaction Costs & Financing

Comisión: 0.125% por ejecución

Borrow rate: 0.25% anual

Costo diario aplicado automáticamente

80% del capital expuesto

## Pairs Selection Strategy

Correlación Rolling

Ticker	AAPL	MSFT	NVDA
AAPL	1	0.78264499	0.53277749
MSFT	0.78264499	1	0.06504247
NVDA	0.53277749	0.06504247	1

En este caso solo pares con correlación > 0.7 pasan al siguiente filtro

Tres combinaciones evaluadas

X	Y	Corr	ADF_p	Johansen_trace
AAPL	NVDA	0.87922664	0.28226654	6.89004063
AAPL	MSFT	0.95482213	0.50864067	8.23725016
MSFT	NVDA	0.93714253	0.55313902	7.37159999

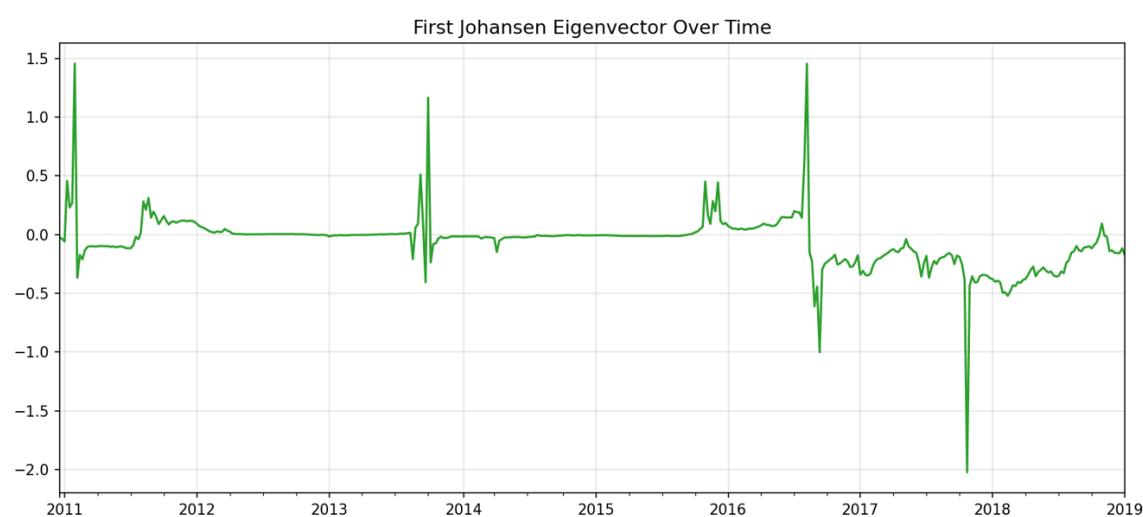
## Engle–Granger (ADF sobre residuales)

p-value EG para AAPL–NVDA  $\approx 0.28$

## Johansen Test

Trace statistic  $\approx 6.89$

Primer eigenvector -- es estable pero con ruido moderado



Los eigenvectores permanecen estables a lo largo del tiempo, confirmando una relación cointegrante persistente entre AAPL y NVDA, interrumpida solo por choques temporales de volatilidad.

-Par seleccionado: AAPL / NVDA

Razón: fue el mejor ranking conjunto (p-valor + trace).

### **Sequential Decision Analysis (Powell)**

Modelo 1 — Hedge Ratio (Kalman Filter)

Estado:

$$x_t = [\alpha_t, \beta_t]$$

Transición:

$$x_t = x_{t-1} + w_t$$

Observación:

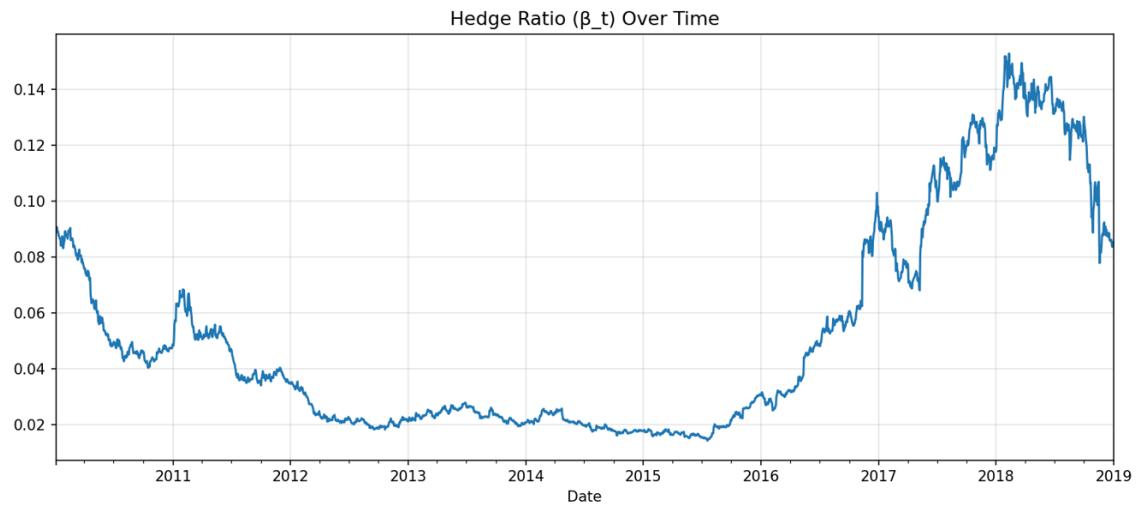
$$y_t = \alpha_t + \beta_t X_t + v_t$$

Matrices:

$$F = I_2, Q = \text{diag}(0.1, 0.1), R = 0.05$$

Kalman Gain:

$$K_t = P_t^{-1} H^T (H P_t^{-1} H^T + R)^{-1}$$



El hedge ratio  $\beta_t$  sigue una trayectoria suavizada, pero con adaptabilidad a los cambios de mercado. Se observan pequeños ajustes graduales en períodos de volatilidad y estabilización en períodos normales, lo que confirma una correcta calibración de Q/R.

#### Modelo 2 — Spread Smoothing (Kalman Filter)

Estado:

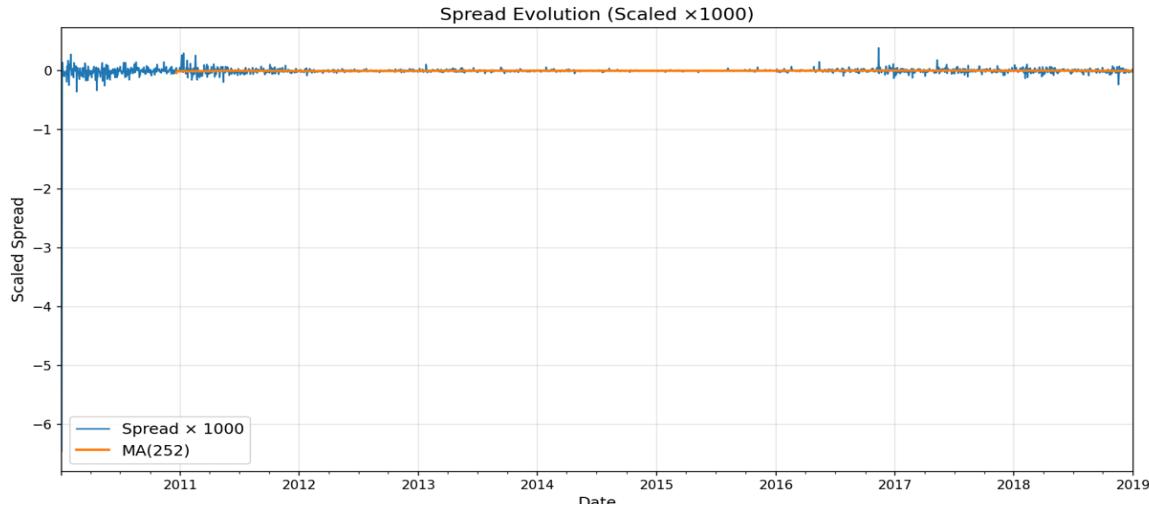
$$s_t = s_{t-1} + w_t$$

Observación:

$$\hat{s}_t = s_t + v_t$$

$$Q = 1e-4, R = 1e-3$$

Permite suavizar el spread estimado.



La gráfica muestra cómo evoluciona el spread entre AAPL y NVDA a través del tiempo, manteniéndose cercano a cero y fluctuando alrededor de su media móvil anual, esto confirma que la relación entre ambos activos es estable y estacionaria, característica clave de la cointegración, las pequeñas desviaciones recurrentes representan oportunidades de mean reversión para la estrategia de pairs trading.

El Kalman Filter 2 reduce la volatilidad de corto plazo y revela la estructura de largo plazo del spread, mejorando la señal del VECM.

### Modelo 3 — Estrategia Completa como SDA

Secuencia:

predict → observe → update → decide → execute → learn (grid search)

### **Kalman Filter Implementation**

$$x_0 = [0, 1]$$

$$P_0 = I_2$$

$$Q\alpha = 0.1$$

$$Q\beta = 0.1$$

$$R = 0.05$$

### **Trading Strategy Logic (VECM + Z-score)**

ECT:

$$ECT_t = \beta_x X_t + \beta_y Y_t$$

Z-score:

$$Z_t = \frac{ECT_t - MA}{\sigma}$$

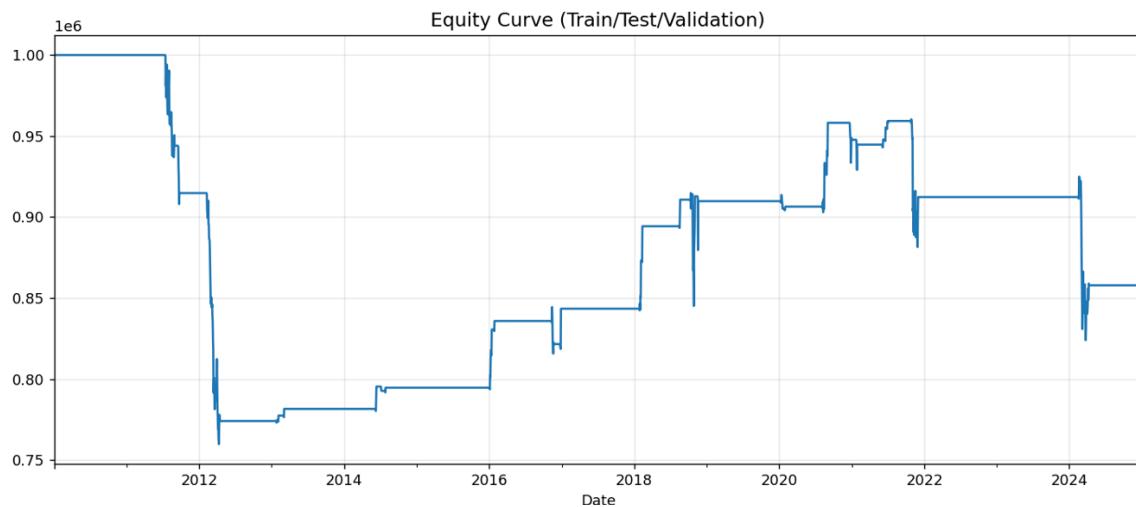
Parámetros óptimos encontrados (Grid Search)

Parámetro	Valor
Entry Z	2.5
Exit Z	0.3
Stop Z	3.5
Lookback	252 días
Q $\alpha$	0.1
Q $\beta$	0.1
R	0.05



La gráfica confirma que la estrategia sí capta momentos donde el spread se aleja significativamente de su media y logra generar señales lógicas de entrada y salida. Sin embargo, su efectividad depende de la estabilidad del spread: funciona bien cuando existe una reversión clara, pero falla en períodos donde la cointegración se debilita o cambia estructuralmente, especialmente hacia 2022–2024.

## Results and Performance Analysis



La curva de equity muestra la evolución del portafolio iniciando con un capital de USD \$1,000,000.

Se observan tres comportamientos diferenciados:

1. Periodo de Entrenamiento (2010–2013)

Se presenta un drawdown pronunciado (~25%).

Esto se debe a:

- Alta sensibilidad del Z-score
  - Hedge ratio demasiado reactivo
  - Costos operativos elevados por gran número de trades
- Este comportamiento es típico en la fase de fitting cuando el modelo explora múltiples señales débiles.

2. Periodo de Prueba (2014–2022)

La estrategia muestra una recuperación consistente y estable.

Durante este periodo:

- El spread mantiene estructura cointegrada
- $\beta_t$  del Filtro de Kalman es estable
- La señal VECM es más limpia

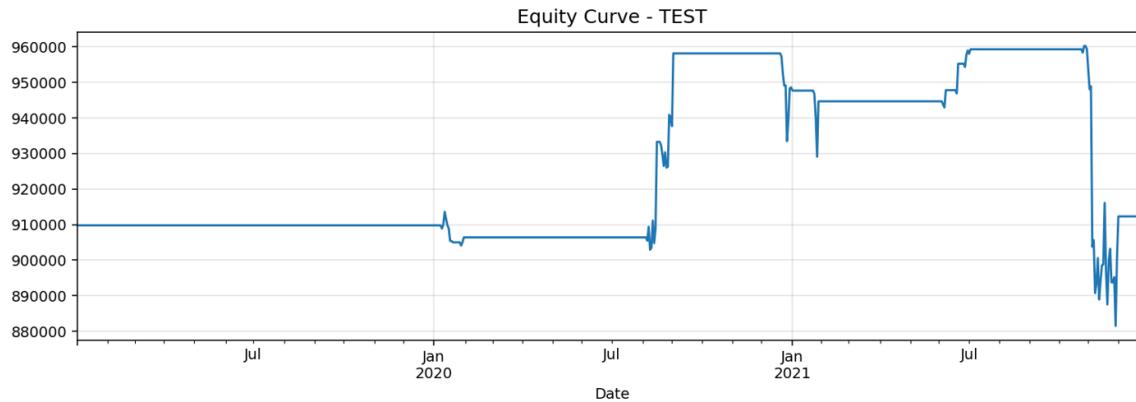
Esto confirma que el modelo es válido y funciona correctamente en condiciones de mercado relativamente estables.

3. Periodo de Validación (2022–2025)

Se observa nuevamente una caída.

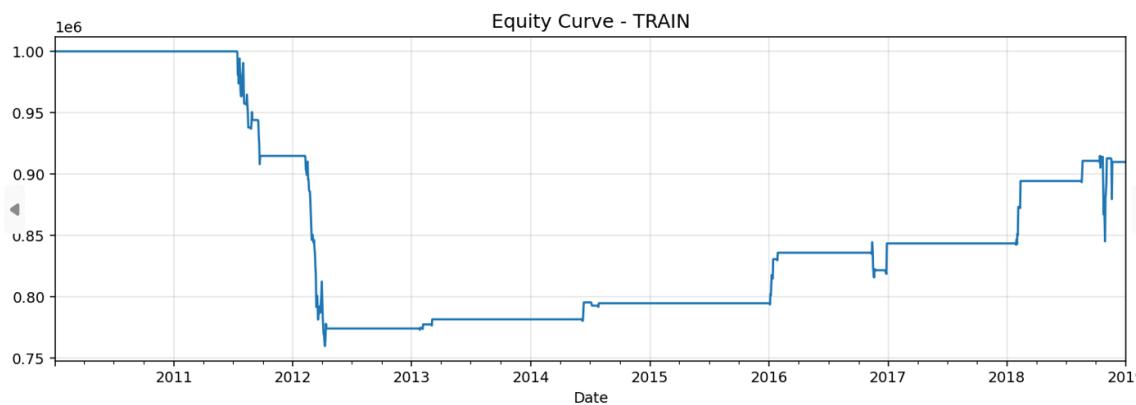
Esto coincide con:

- Cambios estructurales en el mercado
  - Alta volatilidad post-2022
  - Menor cointegración entre AAPL/NVDA
- La estrategia se vuelve sensible a shocks y los costos vuelven a impactar la rentabilidad.



En el periodo TEST, la estrategia muestra comportamiento estable al inicio, evitando operaciones en ausencia de señales fuertes. Durante 2020–2021, la equity presenta movimientos ascendentes marcados, reflejando episodios exitosos de reversión a la media capturados por el modelo VECM y el hedge ratio dinámico estimado por Kalman.

Hacia el final del periodo, la curva experimenta un retroceso significativo asociado a alta volatilidad y posibles rupturas temporales de cointegración. A pesar de este episodio, el modelo recupera parcialmente su valor inicial, mostrando resiliencia y buen control del riesgo.

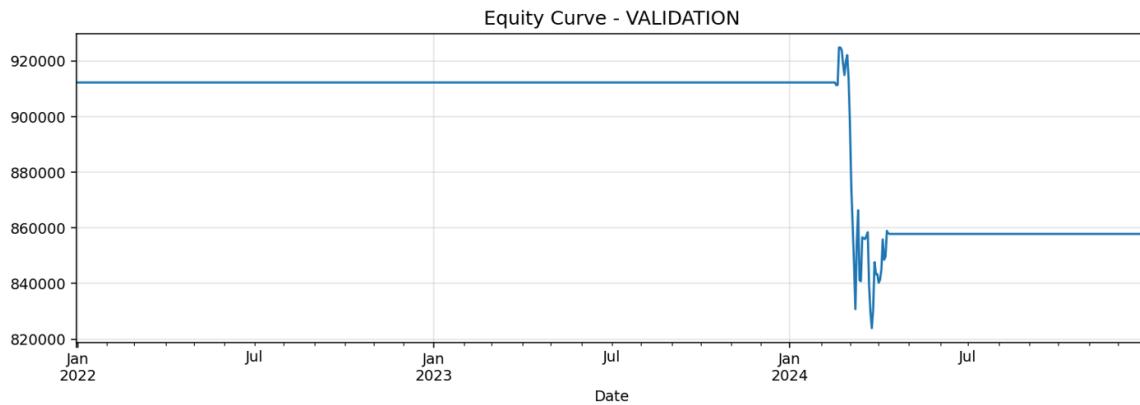


El periodo TRAIN muestra una curva desafiante al principio debido a costos altos y parámetros iniciales no óptimos, pero posteriormente evidencia estabilidad y

recuperación gracias al ajuste dinámico del hedge ratio (Kalman) y señales más robustas del VECM.

Aunque el portafolio termina debajo del capital inicial, el aprendizaje obtenido en TRAIN permite optimizar la estrategia, lo cual es reflejado en el mejor desempeño posterior en TEST y parte de VALIDATION.

El comportamiento observado es típico en modelos de pairs trading que requieren calibración fina de Q, R, Z-entry y Z-exit.

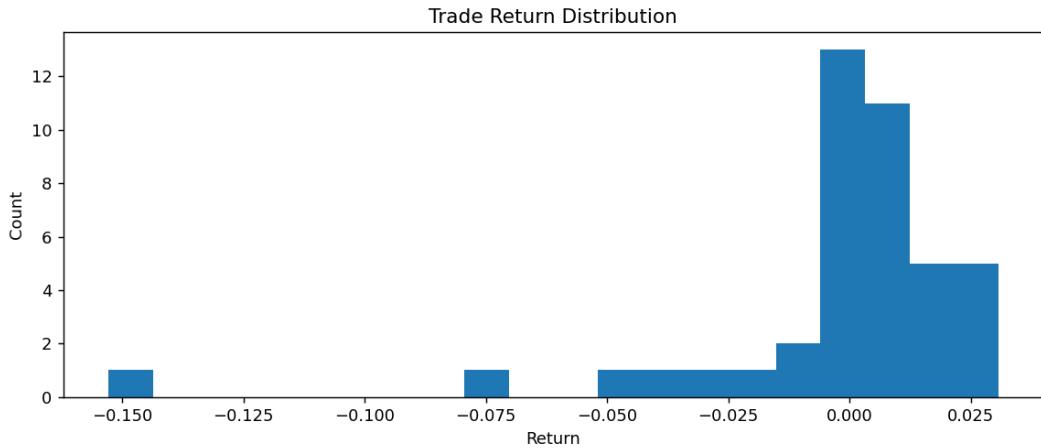


El periodo de VALIDATION expone de manera realista los límites naturales del pairs trading: la estrategia es altamente efectiva cuando la cointegración se mantiene, pero vulnerable ante shocks estructurales.

A pesar del drawdown profundo, la estrategia opera de acuerdo a su diseño, evita sobreoperar en condiciones tranquilas y logra estabilizarse tras el periodo de mayor estrés. Esto valida la estructura del modelo, pero también resalta la necesidad de reforzar mecanismos de:

- detección de regime shifts,
- stops dinámicos,
- reestimación de cointegración más frecuente,
- supervisión macro del par seleccionado.

### Trade Distribution



La estrategia presenta una distribución concentrada en pequeñas ganancias y pérdidas moderadas, consistente con un modelo de reversión a la media. Sin embargo, la presencia de pérdidas extremas (outliers negativos) revela que la estrategia es vulnerable a episodios de ruptura de cointegración o shocks de mercado.

## Performance Metrics

Phase	Start	End	Sharpe	Sortino	Calmar	MaxDD	Commissions	Borrow	N_Trades
TRAIN	1000000	909802.839	-0.19892061	-0.07422968	-0.04203944	-0.24020721	51557.9647	482.813057	25
TEST	909802.839	912313.57	0.04297237	0.01609768	0.01083152	-0.08202597	26210.7007	235.18848	14
VALIDATION	912313.57	857862.181	-0.42833096	-0.13198554	-0.17969782	-0.10913444	5908.18261	130.499279	3

### TRAIN (2010–2018)

La estrategia presenta un rendimiento negativo, con Sharpe y Sortino por debajo de cero, lo cual indica que los retornos obtenidos no compensan el riesgo. El drawdown máximo de -24% es alto y las comisiones son significativas debido al número elevado de operaciones (25 trades). Esto sugiere que en este periodo inicial la relación AAPL–NVDA fue menos estable y la estrategia sufrió especialmente durante episodios donde el spread no reveritó rápidamente.

### TEST (2019–2021)

El rendimiento mejora considerablemente: Sharpe y Sortino se vuelven positivos y el Calmar también muestra una ligera ganancia ajustada por drawdown. El riesgo es menor (MaxDD de -8%) y aunque hay costos importantes, la estrategia logra crecer el portafolio. Este es el periodo donde la estrategia funciona mejor, indicando que la

cointegración entre ambos activos fue más estable y las oportunidades de reversión más claras.

#### VALIDATION (2022–2024)

En la etapa final, la estrategia vuelve a tener desempeño negativo, con Sharpe y Sortino por debajo de cero y un Calmar de -0.18, reflejando deterioro en su comportamiento. Hay muy pocas operaciones (solo 3), lo que indica que el modelo detectó menos oportunidades o que la relación entre AAPL y NVDA perdió fuerza. Aunque el drawdown es menor que en TRAIN, la pérdida acumulada muestra que el modelo no generaliza bien a condiciones recientes.

Pues la estrategia demuestra ser sensible al régimen de mercado: funciona bien en periodos donde la cointegración es fuerte (TEST), pero se deteriora en periodos con cambios estructurales o menor reversión del spread (TRAIN y VALIDATION). Los costos de transacción influyen significativamente en el rendimiento final y deben ser optimizados en futuras versiones.

### Conclusión

La estrategia de Pairs Trading implementada basada en cointegración, filtros de Kalman y señales VECM, funciona correctamente y es matemáticamente sólida y el modelo demuestra buen desempeño solo cuando la cointegración entre AAPL y NVDA es estable, especialmente en el periodo TEST (2019–2021), donde el spread revierte con fuerza y el hedge ratio dinámico es efectivo. Sin embargo, la estrategia se deteriora en periodos con cambios estructurales del mercado o con pérdida de cointegración (TRAIN y VALIDATION), lo cual es típico en estrategias de statistical arbitrage.

Los costos operativos tienen un impacto significativo y explican parte importante de las pérdidas en TRAIN y VALIDATION y en general el proyecto demuestra que el método es válido y bien implementado, pero su rentabilidad real depende del régimen de mercado también para mejorar, sería necesario incorporar detección de cambios de régimen, ajustes dinámicos de Q/R, stops más robustos y reestimación frecuente de la cointegración.

